

LARGE SAMPLE THEORY IN A SEMIPARAMETRIC PARTIALLY LINEAR ERRORS-IN-VARIABLES MODEL

Hua Liang, Wolfgang Härdle and Raymond J. Carroll *

April 4, 1997

Abstract

We consider the partially linear model relating a response Y to predictors (X, T) with mean function $X^T\beta + g(T)$ when the X 's are measured with additive error. The semiparametric likelihood estimate of Severini and Staniswalis (1994) leads to biased estimates of both the parameter β and the function $g(\cdot)$ when measurement error is ignored. We derive a simple modification of their estimator which is a semiparametric version of the usual parametric correction for attenuation. The resulting estimator of β is shown to be consistent and its asymptotic distribution theory is derived. Consistent standard error estimates using sandwich-type ideas are also developed.

Key Words and Phrases: Errors-in-Variables; Functional Relations; Measurement Error; Non-parametric Likelihood; Orthogonal Regression; Partially Linear Model; Semiparametric Models; Structural Relations.

Short title: Partially Linear Models and Measurement Error.

AMS 1991 subject classification: Primary: 62J99, 62H12, 62E25, 62F10 Secondary: 62H25, 62F10, 62F12, 60F05.

*Hua Liang is Associate Professor of Statistics, Institute of Systems Science, Chinese Academy of Sciences, Beijing 100080, China. Wolfgang Härdle is Professor of Econometrics, at the Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, D-10178 Berlin, Germany. Raymond J. Carroll is Professor of Statistics, Nutrition and Toxicology, Department of Statistics, Texas A&M University, College Station, TX 77843-3143. This research was supported in part by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse". The first author was supported by Alexander von Humboldt Foundation. The third author was supported by a grant from the National Cancer Institute (CA 57030) and by the Alexander von Humboldt Foundation. The authors would like to thank Stefan Profit for his comments.

1 INTRODUCTION AND BACKGROUND

Consider the semiparametric partially linear model based on a sample of size n ,

$$Y_i = X_i^T \beta + g(T_i) + \epsilon_i, \quad (1)$$

where X_i is a possibly vector-valued covariate, T_i is a scalar covariate, the function $g(\cdot)$ is unknown, and the model errors ϵ_i are independent with conditional mean zero given the covariates. The partially linear model was introduced by Engle, et al. (1986) to study the effect of weather on electricity demand, and further studied by Heckman (1986), Chen (1988), Speckman (1988), Cuzick (1992a,b), Hua & Härdle (1997) and Severini & Staniswalis (1994).

We are interested in estimation of the unknown parameter β and unknown function $g(\cdot)$ in model (1) when the covariates X are measured with error, and instead of observing X , we observe

$$W_i = X_i + U_i, \quad (2)$$

where the measurement errors U_i are independent and identically distributed, independent of (Y_i, X_i, T_i) , with mean zero and covariance matrix Σ_{uu} . We will assume that Σ_{uu} is known, taking up the case that it is estimated in section 5. The measurement error literature has been surveyed by Fuller (1987) and Carroll, et al. (1995).

If the X 's are observable, estimation of β at ordinary rates of convergence can be obtained by a local-likelihood algorithm, as follows. For every fixed β , let $\hat{g}(T, \beta)$ be an estimator of $g(T)$. For example, in the Severini and Staniswalis implementation, $\hat{g}(T, \beta)$ maximizes a weighted likelihood assuming that the model errors ϵ_i are homoscedastic and normally distributed, with the weights being kernel weights with symmetric kernel density function $K(\cdot)$ and bandwidth h . Having obtained $\hat{g}(T, \beta)$, β is estimated by a least squares operation:

$$\text{minimize } \sum_{i=1}^n \left\{ Y_i - X_i^T \beta - \hat{g}(T_i, \beta) \right\}^2.$$

In this particular case, the estimate for β can be determined explicitly by a projected least squares algorithm. Let $\hat{g}_{y,h}(\cdot)$ and $\hat{g}_{x,h}(\cdot)$ be the kernel regressions with bandwidth h of Y and X on T , respectively. Then

$$\hat{\beta}_x = \left[\sum_{i=1}^n \{X_i - \hat{g}_{x,h}(T_i)\} \{X_i - \hat{g}_{x,h}(T_i)\}^T \right]^{-1} \sum_{i=1}^n \{X_i - \hat{g}_{x,h}(T_i)\} \{Y_i - \hat{g}_{y,h}(T_i)\}. \quad (3)$$

One of the important features of the estimator (3) is that it does not require undersmoothing, so that bandwidths of the usual order $h \sim n^{1/5}$ lead to the result

$$n^{1/2}(\hat{\beta}_n - \beta) \Rightarrow \text{Normal}(0, B^{-1}CB^{-1}), \quad (4)$$

where B is the covariance matrix of $X - E(X|T)$ and C is the covariance matrix of $\epsilon\{X - E(X|T)\}$.

The least squares form of (3) can be used to show that if one ignores measurement error and replaces X by W , the resulting estimate is inconsistent for β . The form though suggests even more. It is well-known that in linear regression, inconsistency caused by measurement error can be overcome by applying the so-called ‘‘correction for attenuation’’. In our context, this suggests that we use the estimator

$$\hat{\beta}_n = \left[\sum_{i=1}^n \{W_i - \hat{g}_{w,h}(T_i)\} \{W_i - \hat{g}_{w,h}(T_i)\}^T - n\Sigma_{uu} \right]^{-1} \sum_{i=1}^n \{W_i - \hat{g}_{w,h}(T_i)\} \{Y_i - \hat{g}_{y,h}(T_i)\}. \quad (5)$$

The estimator (5) can be derived in much the same way as the Severini–Staniswalis estimator. For every β , let $\hat{g}(T, \beta)$ maximize the weighted likelihood ignoring measurement error, and then form β via a negatively penalized operation:

$$\text{minimize } \sum_{i=1}^n \left\{ Y_i - W_i^T \beta - \hat{g}(T_i, \beta) \right\}^2 - \beta^T \Sigma_{uu} \beta. \quad (6)$$

The negative sign in the second term in (6) looks odd until one remembers that the effect of measurement error is attenuation, i.e., to underestimate β in absolute value when it is scalar, and thus one must correct for attenuation by making β larger, not by shrinking it further towards zero.

In this paper, we analyze the estimate (5), showing that it is consistent, asymptotically normally distributed with a variance different from (4). Just as in the Severini–Staniswalis algorithm, in kernel weighting ordinary bandwidths of order $h \sim n^{-1/5}$ may be used.

The outline of the paper is as follows. In Section 2, we define the weighting scheme to be used and hence the estimators of β and $g(\cdot)$. Section 3 is the statement of the main results for β , while the results for $g(\cdot)$ are stated in Section 4. Section 5 states the corresponding results for the measurement error variance Σ_{uu} estimated. Section 6 gives a numerical illustration. Final remarks are given in Section 7. All proofs are delayed until the appendix.

2 DEFINITION OF THE ESTIMATORS

For technical convenience we will assume that the T_i are confined to the interval $[0, 1]$. Throughout, we shall employ $C(0 < C < \infty)$ to denote some constant not depending on n but may assume different values at each appearance. In our proofs and statement of results, we will let the X ’s be unknown fixed constants, a situation which is commonly called the functional relation, see Kendall & Stuart (1992) and Anderson (1984). The results apply immediately to the case that the X ’s are independent random variables, see Section 7.

Let $\omega_{ni}(t) = \omega_{ni}(t; T_1, \dots, T_n)$ be probability weight functions depending only on the design points T_1, \dots, T_n . For example

$$\omega_{ni}(t) = \frac{1}{h_n} \int_{s_{i-1}}^{s_i} K\left(\frac{t-s}{h_n}\right) ds \quad 1 \leq i \leq n \quad (7)$$

where $s_0 = 0, s_n = 1$ and $s_i = (1/2)(T_i + T_{i+1})$, $1 \leq i \leq n-1$. Here h_n is a sequence of bandwidth parameters which tends to zero as $n \rightarrow \infty$ and $K(\cdot)$ is a kernel function, which is supported to have compact support and to satisfy

$$\text{supp}(K) = [-1, 1], \sup |K(x)| \leq C < \infty, \int K(u) du = 1 \text{ and } K(u) = K(-u).$$

In the paper, for any a sequence of variables or functions (S_1, \dots, S_n) , we always denote $\mathbf{S}^T = (S_1, \dots, S_n)$, $\tilde{S}_i = S_i - \sum_{j=1}^n \omega_{nj}(T_i) S_j$, $\tilde{\mathbf{S}}^T = (\tilde{S}_1, \dots, \tilde{S}_n)$. For example, $\tilde{\mathbf{W}}^T = (\tilde{W}_1, \dots, \tilde{W}_n)$, $\tilde{W}_i = W_i - \sum_{j=1}^n \omega_{nj}(T_i) W_j$; $\tilde{g}_i = g(T_i) - \sum_{k=1}^n \omega_{nk}(T_i) g(T_k)$, $\tilde{\mathbf{G}} = (\tilde{g}_1, \dots, \tilde{g}_n)^T$.

The fact that $g(t) = E(Y_i - X_i^T \beta | T = t) = E(Y_i - W_i^T \beta | T = t)$ suggests

$$\hat{g}_n(t) = \sum_{j=1}^n \omega_{nj}(t) (Y_j - W_j^T \hat{\beta}_n) \quad (8)$$

as the estimator of $g(t)$.

In some cases, it may be reasonable to assume that the model errors ϵ_i are homoscedastic with common variance σ^2 . In this event, since $E\{Y_i - X_i^T \beta - g(T_i)\}^2 = \sigma^2$ and $E\{Y_i - W_i^T \beta - g(T_i)\}^2 = E\{Y_i - X_i^T \beta - g(T_i)\}^2 + \beta^T \Sigma_{uu} \beta$, we define

$$\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n (\tilde{Y}_i - \tilde{W}_i^T \hat{\beta}_n)^2 - \hat{\beta}_n^T \Sigma_{uu} \hat{\beta}_n. \quad (9)$$

as the estimator of σ^2 .

3 MAIN RESULTS

We make the following assumptions.

Assumption 1.1. There exist functions $h_j(\cdot)$ defined on $[0, 1]$ such that the j th component of X_i , namely X_{ij} , satisfies $X_{ij} = h_j(T_i) + V_{ij}$, $1 \leq i \leq n$, $1 \leq j \leq p$, where V_{ij} is a sequence of real numbers which satisfy $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n V_i = 0$ and $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n V_i V_i^T = B$ is a positive definite matrix, where $V_i = (V_{i1}, \dots, V_{ip})^T$.

Assumption 1.2. $g(\cdot)$ and $h_j(\cdot)$ are Lipschitz continuous of order 1.

Assumption 1.3. Weight functions $\omega_{ni}(\cdot)$ satisfy:

$$\begin{aligned} (i) \quad & \max_{1 \leq i \leq n} \sum_{j=1}^n \omega_{ni}(T_j) = O(1), \\ (ii) \quad & \max_{1 \leq i, j \leq n} \omega_{ni}(T_j) = O(b_n), \\ (iii) \quad & \max_{1 \leq i \leq n} \sum_{j=1}^n \omega_{nj}(T_i) I(|T_j - T_i| > c_n) = O(c_n), \end{aligned}$$

where $b_n = n^{-4/5}$, $c_n = n^{-1/5} \log n$.

Our two main results concern the limit distributions of the estimate of β and σ^2 .

THEOREM 3.1. Suppose Assumptions 1.1-1.3 hold and $E(\epsilon^4 + \|U\|^4) < \infty$. Then $\hat{\beta}_n$ is an asymptotically normal estimator, i.e.

$$n^{1/2}(\hat{\beta}_n - \beta) \Rightarrow N(0, B^{-1}\Gamma B^{-1}),$$

where

$$\Gamma = E[(\epsilon - U^T \beta)^2 \{X - E(X|T)\} \{X - E(X|T)\}^T] + E\{(UU^T - \Sigma_{uu})\beta\beta^T(UU^T - \Sigma_{uu})\} + E(UU^T \epsilon^2).$$

Note that $\Gamma = E(\epsilon - U^T \beta)^2 B + E\{(UU^T - \Sigma_{uu})\beta\beta^T(UU^T - \Sigma_{uu})\} + \Sigma_{uu}\sigma^2$ if ϵ is homoscedastic and independent of (X, T) .

THEOREM 3.2. Suppose the condition of Theorem 3.1 hold, and that the ϵ 's are homoscedastic with variance σ^2 , and independent of (X, T) . Then

$$n^{1/2}(\hat{\sigma}_n^2 - \sigma^2) \Rightarrow N(0, \sigma_\star^2),$$

where $\sigma_\star^2 = E\{(\epsilon - U^T \beta)^2 - (\beta^T \Sigma_{uu} \beta + \sigma^2)\}^2$.

Remarks

- As described in the introduction, an important aspect of the results of Severini and Staniswalis is that their methods lead to asymptotically normal parameter estimates in kernel regression even with bandwidths of the usual order $h_n \approx n^{-1/5}$. The same holds for our estimators in general. For example, suppose that the design points T_i satisfy that there exist constants M_1, M_2 such that

$$M_1/n \leq \min_{i \leq n} |T_i - T_{i-1}| \leq \max_{i \leq n} |T_i - T_{i-1}| \leq M_2/n.$$

Then Assumptions 1.3(i)-(iii) are satisfied by simple verification.

- It is relatively easy to estimate the covariance matrix of $\hat{\beta}_n$. Let $\dim(X)$ be the number of components of X . A consistent estimate of B is just

$$\{n - \dim(X)\}^{-1} \sum_{i=1}^n \{W_i - \hat{g}_{w,h}(T_i)\} \{W_i - \hat{g}_{w,h}(T_i)\}^T - \Sigma_{uu}.$$

In the general case, one can use (30) to construct a consistent sandwich-type estimate of Γ , namely

$$n^{-1} \sum_{i=1}^n \left\{ \tilde{W}_i(\tilde{Y}_i - \tilde{W}_i^T \hat{\beta}_n) + \Sigma_{uu} \hat{\beta}_n \right\} \left\{ \tilde{W}_i(\tilde{Y}_i - \tilde{W}_i^T \hat{\beta}_n) + \Sigma_{uu} \hat{\beta}_n \right\}^T.$$

In the homoscedastic case, namely that ϵ is independent of (X, T, U) with variance σ^2 , and with U being normally distributed, a different formula can be used. Let $\mathcal{C}(\beta) = E\{(UU^T - \Sigma_{uu})\beta\beta^T(UU^T - \Sigma_{uu})\}$. Then a consistent estimate of Γ is

$$(\hat{\sigma}_n^2 + \hat{\beta}_n^T \Sigma_{uu} \hat{\beta}_n) \hat{B}_n + \hat{\sigma}_n^2 \Sigma_{uu} + \mathcal{C}(\hat{\beta}_n).$$

- In the classical functional model, instead of obtaining an estimate of Σ_{uu} through replication, it is instead assumed that the ratio of Σ_{uu} to σ^2 is known. Without loss of generality, we set this ratio equal to the identity matrix. The resulting analogue of the parametric estimators to the partially linear model is to solve the following minimization problem:

$$\sum_{i=1}^n \left| \frac{\tilde{Y}_i - \tilde{W}_i^T \beta}{\sqrt{1 + \|\beta\|^2}} \right|^2 = \min!$$

where here and in the sequel $\|\cdot\|$ denotes the Euclidean norm. One can use the techniques of this paper to show that this estimator is consistent and asymptotically normally distributed. The asymptotic variance of the estimate of β in this case when ϵ is independent of (X, T) can be shown to equal

$$B^{-1} \left[(1 + \|\beta\|^2) \sigma^2 B + \frac{E\{(\epsilon - U^T \beta)^2 \Gamma_1 \Gamma_1^T\}}{1 + \|\beta\|^2} \right] B^{-1}$$

where $\Gamma_1 = (1 + \|\beta\|^2)U + (\epsilon - U^T \beta)\beta$.

4 ASYMPTOTIC RESULTS FOR THE NONPARAMETRIC PART

THEOREM 4.1. Suppose Assumptions 1.1-1.3 hold and $\omega_{ni}(t)$ are Lipschitz continuous of order 1 for all $i = 1, \dots, n$. If $E(\epsilon^4 + \|U\|^4) < \infty$. Then for fixed T_i , the asymptotic bias and asymptotic variance of $\hat{g}_n(t)$ are respectively, $\sum_{i=1}^n \omega_{ni}(t)g(T_i) - g(t)$ and $\sum_{i=1}^n \omega_{ni}^2(t)(\beta^T \Sigma_{uu} \beta + \sigma^2)$. These are all of order $O(n^{-2/5})$ for kernel estimators.

If the (X_i, T_i) are random, then the bias and variance formulae are the usual ones for nonparametric kernel regression.

5 ESTIMATED ERROR VARIANCE

Although in some cases the measurement error covariance matrix Σ_{uu} has been established by independent experiments, in others it is unknown and must be estimated. The usual method of doing so is by partial replication, so that we observe $W_{ij} = X_i + U_{ij}$, $j = 1, \dots, m_i$.

We consider here only the usual case that $m_i \leq 2$, and assume that a fraction δ of the data has such replicates. Let \bar{W}_i be the sample mean of the replicates. Then a consistent, unbiased method of moments estimate for Σ_{uu} is

$$\hat{\Sigma}_{uu} = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (W_{ij} - \bar{W}_i)(W_{ij} - \bar{W}_i)^T}{\sum_{i=1}^n (m_i - 1)}.$$

The estimator changes only slightly to accommodate the replicates, becoming

$$\begin{aligned} \hat{\beta}_n &= \left[\sum_{i=1}^n \left\{ \bar{W}_i - \hat{g}_{w,h}(T_i) \right\} \left\{ \bar{W}_i - \hat{g}_{w,h}(T_i) \right\}^T - n(1 - \delta/2) \hat{\Sigma}_{uu} \right]^{-1} \\ &\quad \times \sum_{i=1}^n \left\{ \bar{W}_i - \hat{g}_{w,h}(T_i) \right\} \left\{ Y_i - \hat{g}_{y,h}(T_i) \right\}, \end{aligned} \quad (10)$$

where $\hat{g}_{w,h}(\cdot)$ is the kernel regression of the \bar{W}_i 's on T_i .

Using the techniques in the appendix, one can show that the limit distribution of (10) is $\text{Normal}(0, B^{-1} \Gamma_2 B^{-1})$, with

$$\begin{aligned} \Gamma_2 &= (1 - \delta) E \left[(\epsilon - U^T \beta)^2 \{X - E(X|T)\} \{X - E(X|T)\}^T \right] \\ &\quad + \delta E \left[(\epsilon - \bar{U}^T \beta)^2 \{X - E(X|T)\} \{X - E(X|T)\}^T \right] \\ &\quad + (1 - \delta) E \left[\{UU^T - (1 - \delta/2) \Sigma_{uu}\} \beta \beta^T \{UU^T - (1 - \delta/2) \Sigma_{uu}\} + UU^T \epsilon^2 \right] \\ &\quad + \delta E \left[\{\bar{U}\bar{U}^T - (1 - \delta/2) \Sigma_{uu}\} \beta \beta^T \{\bar{U}\bar{U}^T - (1 - \delta/2) \Sigma_{uu}\} + \bar{U}\bar{U}^T \epsilon^2 \right]. \end{aligned} \quad (11)$$

In (11), \bar{U} refers to the mean of two U 's. In the case that ϵ is independent of (X, T) , the sum of the first two terms simplifies to $\{\sigma^2 + \beta^T (1 - \delta/2) \Sigma_{uu} \beta\} B$.

Standard error estimates can also be derived. A consistent estimate of B is

$$\hat{B}_n = \{n - \dim(X)\}^{-1} \sum_{i=1}^n \left\{ \bar{W}_i - \hat{g}_{w,h}(T_i) \right\} \left\{ \bar{W}_i - \hat{g}_{w,h}(T_i) \right\}^T - (1 - \delta/2) \hat{\Sigma}_{uu}.$$

Estimates of Γ_2 are also easily developed. In the homoscedastic case with normal errors, the sum of the first two terms is estimated by $(\hat{\sigma}_n^2 + (1 - \delta/2) \hat{\beta}_n^T \hat{\Sigma}_{uu} \hat{\beta}_n) \hat{B}_n$. The sum of the last two terms is a deterministic function of $(\beta, \sigma^2, \Sigma_{uu})$, and these estimates are simply substituted into the formula.

A general sandwich-type estimator is developed as follows. Define $\kappa = n^{-1} \sum_{i=1}^n m_i^{-1}$, and define

$$R_i = \widetilde{\bar{W}}_i (\tilde{Y}_i - \widetilde{\bar{W}}_i^T \hat{\beta}_n) + \hat{\Sigma}_{uu} \hat{\beta}_n / m_i + (\kappa / \delta) (m_i - 1) \left\{ (1/2) (W_{i1} - W_{i2})(W_{i1} - W_{i2})^T - \hat{\Sigma}_{uu} \right\}.$$

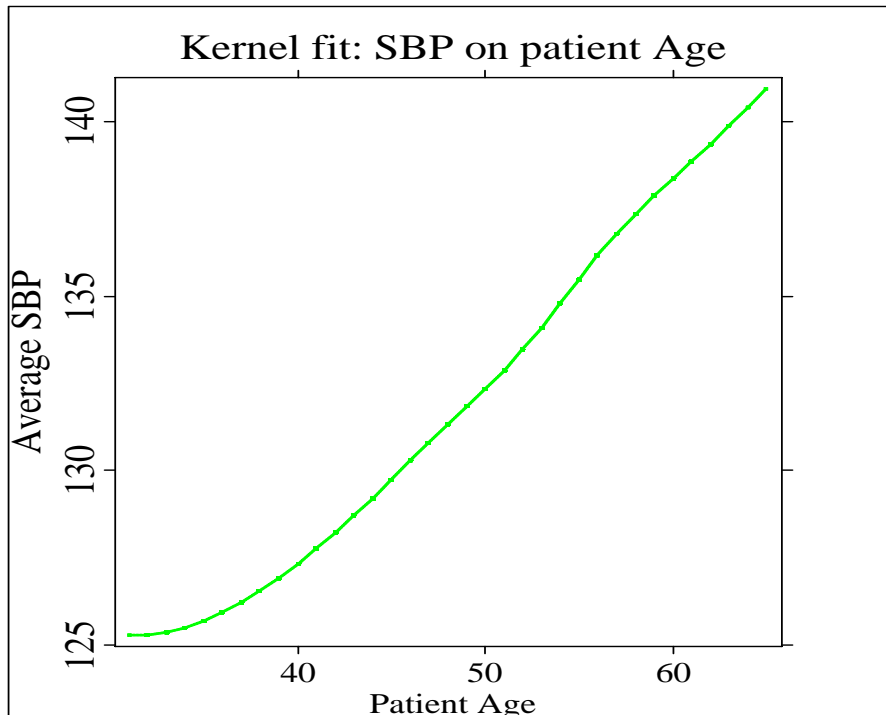


Figure 1: *Estimate of the function $g(T)$ in the Framingham data ignoring measurement error.*

Then a consistent estimate of Γ_2 is the sample covariance matrix of the R_i 's.

6 NUMERICAL EXAMPLE

To illustrate the method, we consider data from the Framingham Heart Study. We considered $n = 1615$ males with Y being their average blood pressure in a fix 2-year period, T being their age and W being the logarithm of the observed cholesterol level, for which there are two replicates.

We did two analyses. In the first, we used both cholesterol measurements, so that in the notation of Section 5, $\delta = 1$. In this analysis, there is not a great deal of measurement error. Thus, in our second analysis, which is given for illustrative purposes, we used only the first cholesterol measurement, but fixed the measurement error variance at the value obtained in the first analysis, in which case $\delta = 0$. For nonparametric fitting, we chose the bandwidth using crossvalidation to predict the response. In precise, we compute the square error using a geometric sequence of 191 bandwidths ranging in $[1, 20]$. The optimal bandwidth is selected to minimize the square error among these 191 candidates. An analysis ignoring measurement error found some curvature in T , see Figure 1 for the estimate of $g(T)$.

As below mention, we will consider four case, using XploRe4 (See Härdle, et al. (1995)) to calculate each case. Our results are as follows. First consider the case that the measurement error was estimated, and both cholesterol values were used to estimate Σ_{uu} . The estimator of β ignoring measurement error was 9.438, with estimated standard error 0.187. When we accounted for measurement error, the estimate increased slightly to $\hat{\beta} = 12.540$, and the standard error increased to 0.195.

In the second analysis, we fixed the measurement error variance and used only the first cholesterol value. The estimator of β ignoring measurement error was 10.744, with estimated standard error 0.492. When we accounted for measurement error, the estimate increased slightly to $\hat{\beta} = 13.690$, and the standard error increased to 0.495.

7 DISCUSSION

Our results have been phrased as if the X 's were fixed constants. If they are random variables, the proofs simplify and the same results are obtained, with now $V_i = X_i - E(X_i|T_i)$.

The nonparametric regression estimator (8) is based on locally weighted averages. In the random X context, the same results apply if (8) is replaced by a locally linear kernel regression estimator.

If we ignore measurement error, the estimator of β is given by (3) but with the unobserved X replaced by the observed W . This differs from the correction for attenuation estimator (5) by a simple factor which is the inverse of the reliability matrix (Gleser, 1992). In other words, the estimator which ignores measurement error is multiplied by the inverse of the reliability matrix to produce a consistent estimate of β . This same algorithm is widely employed in parametric measurement error problems for generalized linear models, where it is often known as an example of regression calibration (see Carroll, et al., 1995, for discussion and references). The use of regression calibration in our semiparametric context thus appears to hold promise when (1) is replaced by a semiparametric generalized linear model.

We have treated the case that the parametric part X of the model has measurement error and the nonparametric part T is measured exactly. An interesting problem is to interchange the roles of X and T , so that the parametric part is measured exactly and the nonparametric part is measured with error, i.e., $E(Y|X, T) = \theta T + g(X)$. Fan and Truong (1993) have shown in this case that with normally distributed measurement error, the nonparametric function $g(\cdot)$ can be estimated only at logarithmic rates, and not with rate $n^{-2/5}$. We conjecture even so that θ is estimable at parametric rates, but this remains an open problem.

REFERENCES

- Anderson, T.W. (1984). Estimating Linear Statistical Relationships. *Annals of Statistics*, 12, 1-45.
- Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995), *Nonlinear Measurement Error Models*. Chapman and Hall, New York.
- Chen, H. (1988). Convergence Rates for Parametric Components in a Partly Linear Model. *Annals of Statistics*, 16, 136-146.
- Cuzick, J. (1992a). Semiparametric Additive Regression. *Journal of the Royal Statistical Society, Series B*, 54, 831-843.
- Cuzick, J. (1992b). Efficient Estimates in Semiparametric Additive Regression Models with Unknown Error Distribution. *Annals of Statistics*, 20, 1129-1136.
- Engle, R. F., Granger, C.W.J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, 81, 310-320.
- Fan, J. & Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Annals of Statistics*, 21, 1900-1925.
- Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York.
- Gleser, L.J. (1992). The importance of assessing measurement reliability in multivariate regression. *Journal of the American Statistical Association*, 87, 696-707.
- Härdle, W., Klinke, S. and Turlach, B.A. (1995). *XploRe: An Interactive Statistical Computing Environment*. Springer-Verlag,
- Heckman, N. E. (1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society, Series B*, 48, 244-248.
- Kendall, M. and Stuart, A. (1992). *The Advanced Theory of Statistics 2*, 4th ed, Charles Griffin, London.
- Liang, H. and Härdle, W. (1997). Asymptotic normality of parametric part in partially linear heteroscedastic regression models. *manuscript*.
- Speckman, P. (1988). Kernel Smoothing in Partial Linear Models. *Journal of the Royal Statistical Society, Series B*, 50, 413-436.
- Severini, T.A. and Staniswalis, J.G. (1994). Quasilikelihood Estimation in Semiparametric Models. *Journal of the American Statistical Association*, 89, 501-511.

8 APPENDIX

In this appendix, we prove several lemmas required. Lemma A.1 provides bounds for $h_j(T_i) - \sum_{k=1}^n \omega_{nk}(T_i)h_j(T_k)$ and $g(T_i) - \sum_{k=1}^n \omega_{nk}(T_i)g(T_k)$. The proof is immediate.

Lemma A.1. Suppose that Assumptions 1.1 and 1.3 (iii) hold. Then

$$\max_{1 \leq i \leq n} |G_j(T_i) - \sum_{k=1}^n \omega_{nk}(T_i) G_j(T_k)| = O(c_n) \quad \text{for } j = 0, \dots, p,$$

where $G_0(\cdot) = g(\cdot)$ and $G_l(\cdot) = h_l(\cdot)$ for $l = 1, \dots, p$.

Lemma A.2. If Assumptions 1.1-1.3 hold. Then

$$\lim_{n \rightarrow \infty} n^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = B$$

Proof. Denote $\bar{h}_{ns}(T_i) = h_s(T_i) - \sum_{k=1}^n \omega_{nk}(T_i) X_{ks}$. It follows from $X_{js} = h_s(T_j) + V_{js}$ that the (s, m) element of $\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ ($s, m = 1, \dots, p$) is

$$\begin{aligned} \sum_{j=1}^n \tilde{X}_{js} \tilde{X}_{jm} &= \sum_{j=1}^n V_{js} V_{jm} + \sum_{j=1}^n \bar{h}_{ns}(T_j) V_{jm} + \sum_{j=1}^n \bar{h}_{nm}(T_j) V_{js} + \sum_{j=1}^n \bar{h}_{ns}(T_j) \bar{h}_{nm}(T_j) \\ &\stackrel{\text{def}}{=} \sum_{j=1}^n V_{js} V_{jm} + \sum_{q=1}^3 R_{nsm}^{(q)} \end{aligned}$$

The strong law of large number implies that $\lim_{n \rightarrow \infty} 1/n \sum_{i=1}^n V_i V_i^T = B$ and Lemma A.1 means $R_{nsm}^{(3)} = o(n)$, which together with the Cauchy-Schwarz inequality show that $R_{nsm}^{(1)} = o(n)$ and $R_{nsm}^{(2)} = o(n)$. This completes the proof of the lemma.

Lemma A.3. (Bennett's inequality) Let $\Gamma_1, \dots, \Gamma_n$ be independent random variables with zero means and bounded ranges: $|\Gamma_i| \leq M$. Then for each $\eta > 0$,

$$P\left\{ \left| \sum_{i=1}^n \Gamma_i \right| > \eta \right\} \leq 2 \exp \left\{ -\eta^2 / \left[2 \left\{ \sum_{i=1}^n \text{var}(\Gamma_i) + M\eta \right\} \right] \right\}.$$

Denote $\epsilon'_j = \epsilon_j I(|\epsilon_j| \leq n^{1/4})$ and $\epsilon''_j = \epsilon_j - \epsilon'_j, j = 1, \dots, n$. We next establish several results for nonparametric regression.

Lemma A.4. Assume that Assumption 1.3 holds and $E\epsilon = 0$ and $E\epsilon^4 < \infty$, Then

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^n \omega_{nk}(T_i) \epsilon_k \right| = O_P(n^{-1/4} \log^{-1/2} n).$$

Proof. Let $M = Cb_n n^{1/4}$. Lemma A.3 implies

$$\begin{aligned} &P \left(\max_{1 \leq i \leq n} \left| \sum_{j=1}^n \omega_{nj}(T_i) (\epsilon'_j - E\epsilon'_j) \right| > C_1 n^{-1/4} \log^{-1/2} n \right) \\ &\leq \sum_{i=1}^n P \left\{ \left| \sum_{j=1}^n \omega_{nj}(T_i) (\epsilon'_j - E\epsilon'_j) \right| > C_1 n^{-1/4} \log^{-1/2} n \right\} \\ &\leq 2n \exp \left\{ -\frac{C_1 n^{-1/2} \log^{-1} n}{\sum_{j=1}^n \omega_{nj}^2(T_i) E\epsilon_j^2 + 2b_n \log^{-1/2} n} \right\} \\ &\leq 2n \exp \{-C_1^2 C \log n\} \rightarrow 0 \quad \text{for some large } C_1 > 0, \end{aligned}$$

which implies that

$$\max_{1 \leq i \leq n} \left| \sum_{j=1}^n \omega_{nj}(T_i)(\epsilon'_j - E\epsilon'_j) \right| = O_P(n^{-1/4} \log^{-1/2} n). \quad (12)$$

On the other hand, we know that

$$\begin{aligned} \max_{1 \leq i \leq n} \sum_{j=1}^n \omega_{nj}(T_i) |E\epsilon''_j| &\leq \max_{1 \leq k \leq n} \max_{1 \leq i \leq n} |\omega_{nk}(T_i)| \sum_{j=1}^n n^{-1} E|\epsilon_j|^4 \\ &\leq C n^{-2/3} E|\epsilon|^4 = O(n^{-1/2}) = O_P(n^{-1/4} \log^{-1/2} n); \end{aligned} \quad (13)$$

$$\sum_{j=1}^n E|\epsilon''_j| \leq n^{-1} \sum_{j=1}^n E\epsilon_j^4 = E\epsilon^4. \quad (14)$$

Moreover, the Hartman-Winter theorem entails that

$$\sum_{j=1}^n (|\epsilon''_j| - E|\epsilon''_j|) = O \left[\left\{ \sum_{j=1}^n E|\epsilon''_j|^2 \log \log \left(\sum_{j=1}^n E|\epsilon''_j|^2 \right) \right\}^{1/2} \right] = O_P(n^{1/4} \log \log^{1/2} n). \quad (15)$$

It follows from (14) and (15) that

$$\sum_{j=1}^n |\epsilon''_j| = O_P(n^{1/4} \log \log^{1/2} n), \quad (16)$$

and

$$\max_{1 \leq i \leq n} \left| \sum_{j=1}^n \omega_{nj}(T_i) \epsilon''_j \right| \leq \max_{1 \leq k, i \leq n} |\omega_{nk}(T_i)| \sum_{j=1}^n |\epsilon''_j| = O(n^{-5/12} \log \log^{1/2} n) = O_P(n^{-1/4} \log^{-1/2} n)$$

Combining the results of (12), (13) with (15), we obtain

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^n \omega_{nk}(T_i) \epsilon_k \right| = O_P(n^{-1/4} \log^{-1/2} n).$$

This completes the proof of Lemma A.4.

Lemma A.5. Assume that Assumptions 1.1-1.3 hold. If $E\epsilon = 0$ and $E\epsilon^4 < \infty$. Then $I_n = o_P(n^{1/2})$, where $I_n = \sum_{i=1}^n \sum_{j \neq i} \omega_{nj}(T_i)(\epsilon'_j - E\epsilon'_j)(\epsilon'_i - E\epsilon'_i)$.

Proof. Let $j_n = \lfloor n^{2/3} \log^2 n \rfloor$, ($[a]$ denotes the integer portion of a .) $A_j = \left\{ \left\lfloor \frac{(j-1)n}{j_n} \right\rfloor + 1, \dots, \left\lfloor \frac{jn}{j_n} \right\rfloor \right\}$, $A_j^c = \{1, 2, \dots, n\} - A_j$ and $A_{ji} = A_j - \{i\}$. Observe that

$$\begin{aligned} I_n &= \sum_{j=1}^{j_n} \sum_{i \in A_j} \sum_{k \in A_{ji}} \omega_{nk}(T_i)(\epsilon'_k - E\epsilon'_k)(\epsilon'_i - E\epsilon'_i) + \sum_{j=1}^{j_n} \sum_{i \in A_j} \sum_{k \in A_j^c} \omega_{nk}(T_i)(\epsilon'_k - E\epsilon'_k)(\epsilon'_i - E\epsilon'_i) \\ &\stackrel{\text{def}}{=} \sum_{j=1}^{j_n} U_{nj} + \sum_{j=1}^{j_n} V_{nj} \stackrel{\text{def}}{=} I_{1n} + I_{2n}, \end{aligned} \quad (17)$$

Where

$$U_{nj} = \sum_{i \in A_j} p_{nij}(\epsilon'_i - E\epsilon'_i) \stackrel{\text{def}}{=} \sum_{i \in A_j} u_{nij}; V_{nj} = \sum_{i \in A_j} q_{nij}(\epsilon'_i - E\epsilon'_i) \stackrel{\text{def}}{=} \sum_{i \in A_j} v_{nij},$$

$$p_{nij} = \sum_{k \in A_{ji}} \omega_{nk}(T_i)(\epsilon'_k - E\epsilon'_k); q_{nij} = \sum_{i \in A_j^c} \omega_{nk}(T_i)(\epsilon'_i - E\epsilon'_i).$$

Notice that $\{v_{nij}, i \in A_j\}$ are conditionally independent random variables given $E_{nj} = \{\epsilon_k, k \in A_j^c\}$ given $E(v_{nij}|E_{nj}) = 0$ $E(v_{nij}^2|E_{nj}) \leq \sigma^2 \max_{1 \leq i \leq n} |q_{nij}|^2 \stackrel{\text{def}}{=} \sigma^2 q_{nj}^2$ for $i \in A_j$, and $\max_{1 \leq i \leq n} |v_{nij}| \leq 2n^{1/4}q_{nj}$, where $q_{nj} = \max_{1 \leq i \leq n} |q_{nij}|$.

On the other hand, by the same reason as (12),

$$q_n = \max_{1 \leq j \leq j_n} |q_{nj}| = \max_{1 \leq j \leq j_n} \max_{1 \leq i \leq n} \left| \sum_{j \in A_j^c} \omega_{nk}(T_i)(\epsilon'_i - E\epsilon'_i) \right| = O_P(n^{-1/4} \log^{-1/2} n)$$

Denote the numbers of the elements in A_j by $\#A_j$. By applying Lemma A.3, we have, for $j = 1, \dots, j_n$,

$$P \left\{ |V_{nj}| > \frac{Cn^{1/2}}{\sqrt{\log nj_n}} \middle| E_{nj} \right\} \leq C \exp \left\{ -\frac{Cn(\log^{-1} n)j_n^{-2}}{\sigma^2 q_n^2 \#A_j + j_n^{-1} n^{1/4} q_n} \right\} \leq Cn^{-2}.$$

It follows by the bounded convergence theorem, the above and $\#A_j \leq \frac{n}{j_n}$ that

$$\begin{aligned} P \left\{ |V_{nj}| > \frac{Cn^{1/2}}{\sqrt{\log nj_n}} \right\} &\leq C \exp \left\{ -\frac{Cj_n^{-2}}{\sigma^2 q_n^2 nj_n^{-1} + j_n^{-1} n^{1/4} q_n} \right\} \\ &\leq Cn^{-2}, \quad \text{for } j = 1, \dots, j_n. \end{aligned}$$

This implies that

$$I_{2n} = o_P(n^{1/2}). \quad (18)$$

Now we consider I_{1n} . Note that $\{\epsilon_k, 1 \leq k \leq n\}$ are i.i.d. random variables, and the definition of U_{nj} , we know that, for any $\zeta > 0$,

$$\begin{aligned} P\{|I_{1n}| > \zeta n^{1/2}\} &\leq \zeta^{-2} n^{-1} E \left(\sum_{j=1}^{j_n} U_{nj} \right)^2 = \zeta^{-2} n^{-1} \left(\sum_{j=1}^{j_n} EU_{nj}^2 + \sum_{j_1 \neq j_2} EU_{nj_1} EU_{nj_2} \right) \\ &\leq \zeta^{-2} n^{-1} j_n (n/j_n)^2 \cdot b_n^2 \left[E(\epsilon' - E\epsilon')^4 + \{E(\epsilon' - E\epsilon')^2\}^2 \right] \\ &\leq \zeta^{-2} n^{-1} \log^{-2} n. \end{aligned} \quad (19)$$

Hence $I_{1n} = o_P(n^{1/2})$. Combining (17), (18) with (19), we complete the proof of Lemma A.5.

Lemma A.6. Assume Assumption 1.3 holds. If $E(|\epsilon|^4 + \|U\|^4) < \infty$. Then

$$\begin{aligned} n^{-1} \sum_{i=1}^n \epsilon_i \left\{ \sum_{j=1}^n \omega_{nj}(T_i) \epsilon_j \right\} &= o_P(n^{-1/2}), \\ n^{-1} \sum_{i=1}^n U_{is} \left\{ \sum_{j=1}^n \omega_{nj}(T_i) U_{jm} \right\} &= o_P(n^{-1/2}), \end{aligned} \quad (20)$$

hold for $1 \leq s, m \leq p$.

Proof. We only prove the first part. The second item is proved similarly. We omit the details.

Using the Hartman-Winter law of the iterated logarithm on $\sum_{k=1}^n (\epsilon_k^2 - E\epsilon_k^2)$, we know that

$$\sum_{k=1}^n \omega_{ni}(T_k) \epsilon_k^2 = o(n^l) \quad \text{for } 0 < l < 1/6 \quad a.s. \quad (21)$$

Observe that

$$\left| \sum_{i=1}^n \sum_{k \neq i}^n \omega_{ni}(T_k) \epsilon_i \epsilon_k \right| \leq \left| \sum_{i=1}^n \sum_{k \neq i}^n \omega_{nk}(T_i) \epsilon_i \epsilon_k - I_n \right| + I_n \stackrel{\text{def}}{=} J_{1n} + I_n.$$

It follows from Lemma A.4, (12), (14) and (16) that

$$\begin{aligned} J_{1n} &\leq \max_{1 \leq j \leq n} \left| \sum_{i=1}^n \omega_{nj}(T_i) \epsilon_i \right| \sum_{i=1}^n (|\epsilon_i''| + E|\epsilon_i''|) \\ &\quad + \max_{1 \leq j \leq n} \left| \sum_{i=1}^n \omega_{nj}(T_i) (\epsilon_i' - E\epsilon_i') \right| \sum_{i=1}^n (|\epsilon_i''| + E|\epsilon_i''|) = o(n^{1/2}). \end{aligned} \quad (22)$$

Therefore, it follows from Lemma A.4 and 21)-(22) that

$$n^{-1} \sum_{i=1}^n \epsilon_i \left\{ \sum_{j=1}^n \omega_{nj}(T_i) \epsilon_j \right\} = n^{-1} \sum_{k=1}^n \omega_{ni}(T_k) \epsilon_k^2 + n^{-1} \sum_{i=1}^n \sum_{k \neq i}^n \omega_{ni}(T_k) \epsilon_i \epsilon_k = o(n^{-1/2}).$$

This completes the proof of (20).

Lemma A.7. Assume that Assumptions 1.1-1.3 holds and $E(\epsilon^4 + \|U\|^4) < \infty$. Then

$$\lim_{n \rightarrow \infty} n^{-1} \widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}} = B + \Sigma_{uu} \quad (23)$$

$$\lim_{n \rightarrow \infty} n^{-1} \widetilde{\mathbf{W}}^T \widetilde{\mathbf{Y}} = B\beta \quad (24)$$

$$\lim_{n \rightarrow \infty} n^{-1} \widetilde{\mathbf{Y}}^T \widetilde{\mathbf{Y}} = \beta^T B \beta + \sigma^2 \quad (25)$$

hold in probability.

Proof. Since $W_i = X_i + U_i$ and $\widetilde{W}_i = \widetilde{X}_i + \widetilde{U}_i$. We get

$$(\widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}})_{sm} = (\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}})_{sm} + (\widetilde{\mathbf{U}}^T \widetilde{\mathbf{X}})_{sm} + (\widetilde{\mathbf{X}}^T \widetilde{\mathbf{U}})_{sm} + (\widetilde{\mathbf{U}}^T \widetilde{\mathbf{U}})_{sm} \quad (26)$$

It follows from law of the strong large number and Lemma A.2 that

$$n^{-1} \sum_{j=1}^n X_{js} U_{jm} \rightarrow 0 \quad a.s. \quad (27)$$

Observe that

$$\begin{aligned} n^{-1} \sum_{j=1}^n \widetilde{X}_{js} \widetilde{U}_{jm} &= n^{-1} \left[\sum_{j=1}^n X_{js} U_{jm} - \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) X_{ks} \right\} U_{jm} \right. \\ &\quad \left. - \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{km} \right\} X_{js} + \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) X_{ks} \right\} \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{km} \right\} \right] \end{aligned}$$

In a fashion similar to Lemma A.4, $\sup_{j \leq n} \sum_{k=1}^n \omega_{nk}(T_j) U_{km} = o_P(1)$, which together with (27) and Assumption 1.3 (ii) entail that the above each term tends to zero. The same reason means that $n^{-1}(\tilde{\mathbf{U}}^T \tilde{\mathbf{X}})_{sm}$ also tends to zero.

Now we prove

$$n^{-1}(\tilde{\mathbf{U}}^T \tilde{\mathbf{U}})_{sm} \rightarrow \sigma_{sm}^2 \quad (28)$$

here σ_{sm}^2 is the (s, m) -th element of Σ_{uu} . Now

$$\begin{aligned} n^{-1}(\tilde{\mathbf{U}}^T \tilde{\mathbf{U}})_{sm} &= n^{-1} \left[\sum_{j=1}^n U_{js} U_{jm} - \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{ks} \right\} U_{jm} \right. \\ &\quad \left. - \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{km} \right\} U_{js} + \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{ks} \right\} \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{km} \right\} \right]. \end{aligned}$$

Obviously $n^{-1} \sum_{j=1}^n U_{js} U_{jm} \rightarrow \sigma_{sm}^2$. Thus from Lemmas A.4, and A.6, (28) holds. Recall (26), (28) and the arguments for $1/n(\tilde{\mathbf{U}}^T \tilde{\mathbf{X}})_{sm} \rightarrow 0$ and $1/n(\tilde{\mathbf{X}}^T \tilde{\mathbf{U}})_{sm} \rightarrow 0$, we complete the proof of (23).

Next we prove (24). Note that $\tilde{\mathbf{W}}^T \tilde{\mathbf{Y}} = \tilde{\mathbf{W}}^T (\tilde{\mathbf{X}}\beta + \tilde{\mathbf{G}} + \tilde{\epsilon})$. This follows because

$$\left| \sum_{j=1}^n X_{js} \tilde{g}_j \right| \leq \left(\sum_{j=1}^n X_{js}^2 \sum_{j=1}^n \tilde{g}_j^2 \right)^{1/2} \leq c_n n^{1/2} \left(\sum_{j=1}^n X_{js}^2 \right)^{1/2} \leq C n c_n,$$

and

$$\begin{aligned} (\tilde{\mathbf{W}}^T \tilde{\mathbf{G}})_s &= \sum_{j=1}^n \tilde{X}_{js} \tilde{g}_j + \sum_{j=1}^n \tilde{U}_{js} \tilde{g}_j \\ &= \sum_{j=1}^n \left\{ X_{js} - \sum_{k=1}^n \omega_{nk}(T_j) X_{ks} \right\} \tilde{g}_j + \sum_{j=1}^n \tilde{U}_{js} \tilde{g}_j. \end{aligned}$$

Obviously $1/n \sum_{j=1}^n \tilde{U}_{js} \tilde{g}_j$ tends to zero. Therefore $1/n(\tilde{\mathbf{W}}^T \tilde{\mathbf{G}})_s$ tends to zero.

The proof of $1/n(\tilde{\mathbf{W}}^T \tilde{\epsilon})_s$ tends to zero is similar that of $1/n(\tilde{\mathbf{W}}^T \tilde{\mathbf{U}})_s \rightarrow 0$. Combining the above arguments and (23), we complete the proof (24). The proof of (25) can be completed by the above similar arguments. The details are omitted.

Proof of Theorem 3.1. Denote $\Delta_n = (\tilde{\mathbf{W}}^T \tilde{\mathbf{W}} - n\Sigma_{uu})/n$. By Lemma A.7 and direct calculation,

$$\begin{aligned} n^{1/2}(\hat{\beta}_n - \beta) &= n^{-1/2} \Delta_n^{-1} (\tilde{\mathbf{W}}^T \tilde{\mathbf{Y}} - \tilde{\mathbf{W}}^T \tilde{\mathbf{W}} \beta + n\Sigma_{uu} \beta) \\ &= n^{-1/2} \Delta_n^{-1} (\tilde{\mathbf{X}}^T \tilde{\mathbf{G}} + \tilde{\mathbf{X}}^T \tilde{\epsilon} + \tilde{\mathbf{U}}^T \tilde{\mathbf{G}} + \tilde{\mathbf{U}}^T \tilde{\epsilon} - \tilde{\mathbf{X}}^T \tilde{\mathbf{U}} \beta - \tilde{\mathbf{U}}^T \tilde{\mathbf{U}} \beta + n\Sigma_{uu} \beta). \end{aligned}$$

By Lemmas A.2, A.4–A.6, we conclude that

$$n^{1/2}(\hat{\beta}_n - \beta) = n^{-1/2} \Delta_n^{-1} \sum_{i=1}^n (V_i \epsilon_i - V_i U_i^T \beta + U_i \epsilon_i - U_i U_i^T \beta + \Sigma_{uu} \beta) + o_P(1) \quad (29)$$

$$\stackrel{\text{def}}{=} n^{-1/2} \sum_{i=1}^n \zeta_{in} + o_P(1). \quad (30)$$

Although the $\{V_i\}$ are nonrandom variables, because

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n V_i = 0; \quad \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n V_i V_i^T = B \quad (31)$$

and $E(\epsilon^4 + \|U\|^4) < \infty$, it follows that the sequence of k -th elements $\{\zeta_{in}^{(k)}\}$ of $\{\zeta_{in}\}$ ($k = 1, \dots, p$) satisfy, for any given $\zeta > 0$, $n^{-1} \sum_{i=1}^n E\{\zeta_{in}^{(k)2} I(|\zeta_{in}^{(k)}| > \zeta n^{1/2})\} \rightarrow 0$ as $n \rightarrow \infty$. This means that Lindeberg's condition for the central limit theorem holds. Moreover, note that

$$\begin{aligned} Cov(\zeta_{ni}) &= E\{V_i(\epsilon_i - U_i^T \beta)^2 V_i^T\} + E\{(U_i U_i^T - \Sigma_{uu})\beta \beta^T (U_i U_i^T - \Sigma_{uu})\} + E(U_i U_i^T \epsilon_i^2) \\ &\quad + V_i E(U_i^T \beta \beta^T U_i U_i^T) + E(U_i U_i^T \beta \beta^T U_i) V_i \end{aligned}$$

which and (31) entail that

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n Cov(\zeta_{ni}) = E(\epsilon - U^T \beta)^2 B + E\{(U \cdot U^T - \Sigma_{uu})\beta \beta^T (U \cdot U^T - \Sigma_{uu})\} + E(U U^T \epsilon^2).$$

Theorem 3.1 now follows.

Proof of Theorem 3.2. Denote

$$A_n = n^{-1} \begin{bmatrix} \tilde{\mathbf{Y}}^T \tilde{\mathbf{Y}} & \tilde{\mathbf{Y}}^T \tilde{\mathbf{W}} \\ \tilde{\mathbf{W}}^T \tilde{\mathbf{Y}} & \tilde{\mathbf{W}}^T \tilde{\mathbf{W}} \end{bmatrix}; \quad A = \begin{bmatrix} \beta^T B \beta + \sigma^2 & \beta^T B \\ B \beta & B + \Sigma_{uu} \end{bmatrix};$$

$$\tilde{A}_n = n^{-1} \begin{bmatrix} (\epsilon + U \beta)^T (\epsilon + U \beta) & (\epsilon + U \beta)^T (U + V) \\ (U + V)^T (\epsilon + U \beta) & (U + V)^T (U + V) \end{bmatrix}.$$

According to the definition of $\hat{\sigma}_n^2$, direct calculation using Lemma A.6 yields that

$$n^{1/2}(\hat{\sigma}_n^2 - \sigma^2) = n^{1/2} \left(\sum_{j=1}^5 S_{jn} \right) + \frac{1}{n^{1/2}} (\epsilon - \tilde{\mathbf{U}} \beta)^T (\epsilon - \tilde{\mathbf{U}} \beta) - n^{1/2} (\hat{\beta}_n^T \Sigma_{uu} \hat{\beta}_n + \sigma^2) + o_P(1),$$

where

$$\begin{aligned} S_{1n} &= (1, -\hat{\beta}_n^T) (A_n - \tilde{A}_n) (1, -\hat{\beta}_n^T)^T \\ S_{2n} &= (1, -\hat{\beta}_n^T) (\tilde{A}_n - A) (0, \beta^T - \hat{\beta}_n^T)^T \\ S_{3n} &= (0, \beta^T - \hat{\beta}_n^T) A (0, \beta^T - \hat{\beta}_n^T)^T \\ S_{4n} &= (0, \beta^T - \hat{\beta}_n^T) (\tilde{A}_n - A) (1, -\beta^T)^T \\ S_{5n} &= -(\beta - \hat{\beta}_n)^T (\beta - \hat{\beta}_n). \end{aligned}$$

It follows from Theorem 2.1 and Lemma A.7, $n^{1/2} \sum_{j=1}^5 S_{jn} \rightarrow 0$ in probability, and $n^{1/2}(\hat{\sigma}_n^2 - \sigma^2) = \frac{1}{n^{1/2}} \sum_{i=1}^n \{(\epsilon_i - U_i^T \beta)^2 - (\beta^T \Sigma_{uu} \beta + \sigma^2)\} + o_P(1)$. Theorem 3.2 now follows immediately.

Proof of Theorem 4.1. Since $\hat{\beta}_n$ is a consistent estimator of β , its asymptotic bias and variance equal the relative ones of $\sum_{j=1}^n \omega_{nj}(t)(Y_j - W_j^T \beta)$, which is denoted by $\hat{g}_n^*(t)$. By simple calculation,

$$\begin{aligned} E\hat{g}_n^*(t) - g(t) &= \sum_{i=1}^n \omega_{ni}(t)g(T_i) - g(t), \\ \hat{g}_n^*(t) - E\hat{g}_n^*(t) &= \sum_{i=1}^n \omega_{ni}^2(t)(\beta^T \Sigma_{uu} \beta + \sigma^2). \end{aligned}$$

Both of them are order $O(n^{-1/5})$ by Lemma A.1 and Assumption 1.3 (iii). Theorem 4.1 is immediately proved.